Karakostas A, Zizzo DJ.

Compliance and the Power of Authority.

***Journal of Economic Behavior & Organization* (2015)**

http://dx.doi.org/10.1016/j.jebo.2015.09.016

Journal of Economic Behavior & Organization xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Compliance and the power of authority

Alexandros Karakostas[a,**], Daniel John Zizzo[b,*]

[a] University of Erlangen-Nuremberg, Germany
[b] Newcastle University, United Kingdom

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form
18 September 2015
Accepted 21 September 2015
Available online xxx

JEL classification:
C72
C91
D03
M12
H8

Keywords:
Compliance
Obedience
Conformism
Social pressure
Experimenter demand
Antisocial behavior

ABSTRACT

Compliance to authority is an integral part of how organizations operate. We use an experiment to show that compliance to a cue by an authority is a powerful motivating mechanism. We do this in an experiment where there are direct orders or indirect cues to destroy half of another participant's earnings at a cost to one's own earnings. Depending on the experimental treatment, up to around 60–70% of participants decide to comply with the orders or cues being provided.

> Within the scope of the contract, the relation between employer and employee is no longer a market relation but an authority relation.
>
> (Arrow, 1974:64)

1. Introduction

Kenneth Arrow (1974) argued that the most prevalent characteristic of organizations of any size is the relationship of authority and that this, along with compliance to authority, is an integral part of the mechanism by which organizations function. Organizational psychologists such as Cialdini and Goldstein (2004, p. 596) claimed that "most organizations would cease to operate efficiently if deference to authority were not one of the prevailing norms." In social psychology, the paradigmatic case is provided by Milgram's (1963, 1974) series of psychology experiments on authority. In these studies the subjects, for the sake of science, were requested to press a button supposedly to implement an escalating series of electricity shocks

* Corresponding author at: BENC and Newcastle University Business School, Newcastle University, 5 Barrack Road, Newcastle upon Tyne, Tyne and Wear NE1 4SE, United Kingdom. Tel.: +44 191 2085844.
** Corresponding author.
E-mail addresses: alexandros.karakostas@fau.de (A. Karakostas), daniel.zizzo@ncl.ac.uk (D.J. Zizzo).

http://dx.doi.org/10.1016/j.jebo.2015.09.016

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

to a confederate every time the latter provided a wrong answer. The results were surprising, as up to 62.5% of the subjects continued up to the maximum 450 volts electric shock.[1] These studies involved deception and a strong authority presence who provided progressively stronger verbal cues as the experiment progressed. In economics, authority has been examined in terms of decision rights and asymmetric information (Grossman and Hart, 1986; Hart and Moore, 1990; Aghion and Tirole, 1997). Yet, the psychological effect of authority on economic interactions has been largely disregarded.

A few recent experimental studies have hinted at the potential importance of the psychological effect of authority for tax compliance. Silverman et al. (2014) presented a public good experiment where subjects were requested, and later on reinforced with appeal to the expertise of different authorities,[2] to contribute the socially optimal level. An explanation for why subjects should behave prosocially tended to be effective when combined with the threat of punishment (Silverman et al., 2014).[3] Similarly, Cadsby et al. (2006) have shown that many people pay taxes because they are inclined to obey authority despite the financial attraction of non-compliance. Cadsby et al. (2006) comes closest to our experiment by requiring subjects to pay a cost. However, their design is set in a prosocial context[4] and in their manipulation subjects were not simply requested but also *expected* to act in line with the request of the experimenter's authority.[5] In an organizational frame, Robin et al. (2012) found that workers are willing to change their opinions in order to comply with those of their managers.

We define compliance to authority as a willingness to permit one's behavior to be determined by the experimenter.[6] We present a simple experiment on the role of compliance towards an authority, the experimenter, in an experiment where compliance damages other people's earnings at a small own monetary cost. In our experiment we examine how (in)direct cues for compliance affect participants' willingness to comply. Additionally, we examine the implications of these cues being: (un)justified, in the form of an order, or provided at specific intervals in time, affect participants' compliance. In that sense our experiment also serves as an experimental study on 'nudging' (see Thaler and Sunstein, 2008) and particularly how different nudges affect participants' compliance to requests by an authority. There is significant experimental evidence which suggests that some people pay taxes because they are intrinsically motivated to pay taxes or more generally to obey the law (see, e.g., Feld and Frey, 2002 or Alm and Torgler, 2006). In these studies paying tax is framed as a moral obligation. In our design there are neither prosocial incentives towards the other participants (as in Silverman et al., 2014), nor monetary incentives for compliance with the authority (as in Robin et al., 2012). This allows us to focus on whether participants have a drive to comply with an authority, even when there are no monetary benefits (or the risk of sanctions) for him or his peers from doing so. This enables us to shed more light on whether a norm for compliance exists and if so how it influences behavior. In this respect, the choice of an antisocial domain for the action to be undertaken by the agent is especially useful to control for potential explanations of changes in behavior based on social norms to be prosocial (e.g., Bicchieri, 2006; Keizer et al., 2008; Bicchieri and Xiao, 2009).[7] Within a tax context, this is in turn relevant to identify possible strategies to increase tax payment, for example based on advertising campaigns.

A laboratory experiment is particularly useful to understand *whether* there is such a thing as deference to authority and *why* it exists, i.e., to identify the underlying preferences and mechanisms that would be difficult to isolate with non-laboratory data. For example, outside the laboratory, there are often long term relationships between an authority and the subordinate. In the workplace, for instance, it is difficult to cleanly identify whether and why there is deference to authority as opposed to simply behaving in a way that satisfies one's own long-term reputational self-interest. There is a natural way to implement authority in the laboratory, and that is by exploiting the authority of the experimenter: subjects view the experimenter as being in a position of authority due to its legitimacy and expertise about the experimental environment (Orne, 1962; Rosnow and Rosenthal, 1997; Zizzo, 2010). Similarly to Cadsby et al. (2006) and Silverman et al. (2014), we use experimenter demand as our tool to study compliance.[8]

We wish to address the following questions in this paper. Can *vertical*, i.e., hierarchical, social pressure by an authority induce compliance? What if there is no financial incentive for complying, the action to be undertaken by the agent is antisocial (that is, contrary to the standard social norm of not damaging others), and the nudge is much weaker than in Milgram's

[1] Similar results have been found under a range of variants and of subject pools (e.g., Kilham and Mann, 1974; Shanab and Yahya, 1977, 1978; Meeus and Raaijmakers, 1995; Blass, 1999). See Reicher et al. (2012) for a recent review and reinterpretation of the experimental psychology literature.

[2] Initially there is an explicit request, later on reinforced with a justification provided by four graduate students, or four members of faculty.

[3] Arrow (1974) also argued that the main role for sanctions is to facilitate authority, an argument strongly supported by the findings by Silverman et al. (2014).

[4] For reasons explained below, we preferred an antisocial context, as it allows us to eliminate potential confounds.

[5] Subjects were given an explicit reason to give, which was to help fund future experiments.

[6] Our definition of authority is very similar to Simon's (1979) who defined obedience to authority as "a willingness to permit one's behavior to be determined by the employer" (Simon, 1979:502). Similarly to Arrow (1974), Simon (1946, 1979) also argued that obedience to authority is a prevailing norm within organizations.

[7] Hence, our choice for an antisocial domain is methodological: we wanted to eliminate compliance driven by a motive to be nice towards peers (i.e., the other participants). Nevertheless, there are cases in the real world where subordinates need to take unpleasant or nasty decisions driven by requests from a higher authority. Two examples include mid-level managers who need to reduce personnel, or alternatively relocate staff to shared offices, under a directive from a high-level manager.

[8] Providing the role of authority to the experimenter is a considerably more valid experimental test of the role of the authority in the laboratory than would, say, providing the role of authority to an experimental subject; the latter would make the experiment one where any effect of authority would be inexorably confounded with one of peer pressure, rather than enabling to identify the effect of authority.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

experiments? And if so, does providing a justification for acting antisocially matters, and can we induce greater compliance by requesting to comply on specific intervals in time?

Overall, our results suggest that greater attention on the psychological effect of authority should be given in principal-agent modeling and more generally in thinking about incentives and delegation in public and private organizations. We find that an indirect nudge to comply with our request to destroy half of another's participant earnings does not make any statistical difference when compared to a treatment that no cue is provided. However, when indirect pressure is reinforced by an explicit request in the instructions to reduce their partners' income in the form of a polite request with respect to the action being 'useful', even if there is no explicit reason provided for the usefulness, the destruction rates more than doubles. If the request is phrased as an order, the increase is not as large. Asking subjects to destroy in specific intervals in time increases destruction rates further, a result robust to providing an explicit justification for destruction. Additionally, we find that subjects are willing to destroy half of their partners' earnings from the experiment even if their partner is unable to destroy their own earnings (i.e., destruction due to the potential for reciprocal aggression is ruled out), and beliefs about partner behavior can only partly explain compliance to authority.

Obviously, the way that we frame our request may matter, and further research needs to look beyond what we are able to do in this paper. Similarly, while we follow Milgram in treating the experimenter as the authority, there is an obvious question of how generalizable this is, which is also left for future research.

The remainder of this paper is structured as follows. Section 2 considers some further related literature. Section 3 provides the experimental design and research hypotheses. Section 4 presents the results, and Section 5 the discussion and conclusions.

2. Further related literature

Principal-agent models are the standard conceptual framework by which economists consider authority, but their focus has been on analyzing the effect of different pay structures on decision making by agents. While many experiments have studied the impact of social preferences on employment contracts (e.g., Anderhub et al., 2002; Fehr et al., 2007; Karakostas et al., 2013), consideration has not been given to how authority per se may help induce compliance to the employer's expected effort; and, while one can have experiments in which the principal suggests effort (as in Güth et al., 2001), students are likely to see other students as peers rather than authorities. There has been considerable attention in economics to conformism and social norms, by which subjects tend to do what a number of others do (e.g., Asch, 1955; Jones, 1984; Lopez-Perez, 2008; Zafar, 2011), and there is an important empirical literature on peer effects (e.g., Case and Katz, 1991; Kawaguchi, 2004; Powell et al., 2005; Lundborg, 2006) and on social image and prosocial behavior (e.g., Glazer and Konrad, 1996; Benabou and Tirole, 2006; Andreoni and Bernheim, 2009; Ariely et al., 2009), but the focus of this research has been on what we may label as *horizontal* social pressure, i.e., pressure by peers.[9] There has also been some insightful attention to the study of leadership as a way of helping to solve social dilemmas or weakest link type coordination problems (Moxnes and van der Heijden, 2003; Brandts et al., 2007; Güth et al., 2007; Van der Heijden et al., 2008), but the leaders in this literature are just peers whose actions may facilitate cooperation and contribution within groups.

There are a number of other connected strands of research. One is about experimenter demand characteristics by which a number of experiments are criticized on the grounds that experimental subjects may change their behavior due to implicit cues about what constitutes appropriate behavior (Orne, 1962; Rosnow and Rosenthal, 1997; Zizzo, 2010; Zizzo and Fleming, 2011). Other related literature is where information is provided on others' behavior, e.g., in the context of giving (Cason and Mui, 1998; Frey and Meier, 2004; Landry et al., 2006; Krupka and Weber, 2009; Shang and Croson, 2009; Bicchieri and Xiao, 2009) or of public good contribution (Bardsley and Sausgruber, 2005), or where there are exogenous recommendations given in contexts where it can serve one's own self-interest, such as threshold public good games (Croson and Marks, 2001), Cournot markets (Sonntag and Zizzo, 2015) or as a device to solve coordination in Chicken games (Cason and Sharma, 2007; Duffy and Feltovich, 2008; Bone et al., 2012) or to improve effort in gift exchange games (Thöni and Gächter, 2012).

Counter to the usual emphasis of behavioral economists on cooperation, there is a small but growing body of research on antisocial behavior to which this paper is also related to.[10] This research finds that there are conditions under which subjects are willing to pay money to damage other's earnings, even at a cost to their own (for examples, see Zizzo and Oswald, 2001; Zizzo, 2003; Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011; Abbink et al., 2010; Engel and Zhurakhovska, 2012).[11] There are also conditions in which 'antisocial' punishment takes place (Nikiforakis, 2008; Nikiforakis and Engelmann, 2008; Herrmann et al., 2008; Denant-Boemont et al., 2007).

[9] The difficulties with identification in peer effects estimations imply that this is an area where experimental research is especially useful (e.g., Manski, 1993). In psychology, the connection is with traditional research on the value of conformism to peers (e.g., Asch, 1955) and on influence (e.g., Cialdini and Goldstein, 2004).

[10] Herrmann and Orzen (2008) refer to this as *homo rivalis* behavior.

[11] Abbink and Herrmann's (2011) idea that subjects are more willing to engage in antisocial behavior if the moral cost of it is lower has some parallelism with the 'moral wriggle room' literature (e.g., Dana et al., 2007).

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

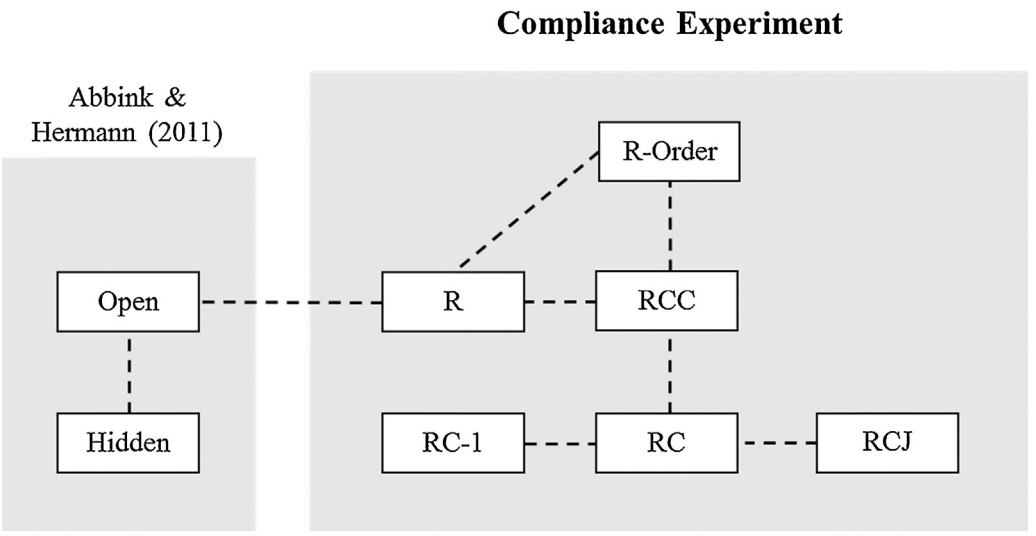

Fig. 1. Experimental design. *Notes.* Each box represents an experimental treatment. R stands for Repeated, RC for Repeated Compliance, RCC for Repeated Compliance Constant pressure, RC-1 for Repeated Compliance 1-sided and RCJ for Repeated Compliance Justified.

3. Experimental design and hypotheses

3.1. Outline

The experiment was conducted at the University of East Anglia between January and March 2012 with 390 subjects. The participants were mostly undergraduate students (79%) while the remainder were postgraduate students. There were participants from 36 different nationalities while the majority of the subjects were British (54%). The mean age was 22 years old and 14.62% of the participants were economics students.[12] The experiment was in paper and pencil.[13] The instructions were as close as possible to those of Abbink and Herrmann (2011). The experiment employed a fictional currency, called *Guilders*, which was converted to pounds at the end of the experiment at the rate of £0.75 per Guilder. Each session lasted approximately 60 min and the subjects earned on average £8.14 (approximately 12.38 US dollars), including a show-up fee of £2.00. Earnings were paid privately and anonymously at the end of the experiment. Subjects were not allowed to participate in more than one session. The experiment consisted of eight treatments (see Fig. 1), described below: Open (O), Hidden (H), Repeated question (R), Repeated Order (R-Order), Repeated Compliance Constant pressure (RCC), Repeated Compliance (RO), Repeated Compliance Justified (RCJ), and Repeated Compliance 1-sided (RC-1). We ran 20 sessions in total.[14]

3.2. The Open and Hidden treatments (O and H)

We began with a straightforward replication of Abbink and Herrmann's (2011) two treatments (O and H). The O treatment employs the *Joy of Destruction game* in which two players were endowed with 10 Guilders each, and both players simultaneously decided whether or not to destroy 5 Guilders of the other player's endowment, at an own cost of 1 Guilder. If they both reduced their partners' income, they both earned 4 Guilders. If one reduced his/her partner's income but the other did not, the first earned 9 Guilders and the second earned 5 Guilders (and vice versa). If no one reduced the other person's income they both earned 10 Guilders.

In the H treatment, a die was also rolled for each player. If it turned out 1 or 6 (1/3 probability) the player would lose 5 Guilders regardless of the other player's decision; if the die turned out 2, 3, 4, or 5 (2/3 probability) the partner's decision was implemented. A player who lost 5 Guilders was not told whether this was due to the partner's action, or because of the roll of the die. Abbink and Herrmann (2011) predicted that in the H treatment the destruction rate would be higher than in the O treatment, as the moral costs of nastiness would be decreased by the player being able to hide behind the possibility of destruction by nature and being able to reason that, had the player not destroyed, destruction may have occurred anyway.

[12] We controlled for economic students in our regression analysis but we did not find it to have a statistically significant impact.
[13] We decided to run this experiment in paper and pencil deliberately in order to be as close as possible to the design of Abbink and Herrmann (2011).
[14] We aimed for (at least) 40 independent observations per treatment, which meant 40 subjects for the O, H, R and R-Order treatments, and 80 subjects for the RC-1 treatment (since, as discussed later, only half of the RC-1 subjects made actual destruction decisions); as we were able to have a few more subjects, we had 56 subjects for the RC treatment and 54 for the RCC treatment.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

Following Abbink and Herrmann (2011), after subjects had decided if they want to reduce their partner's income or not, we used an incentivized questionnaire in which we asked the participants about their beliefs of their partner's behavior (i.e., their choice whether to destroy or not). If their prediction was correct they were rewarded with 1 Guilder.

After the incentivized questionnaire and before subjects were informed about their earnings we requested them to complete two questionnaires. The first was a social desirability scale questionnaire (Stöber, 2001) and the second collected demographic information.

3.3. The Repeated question treatment (R)

The Repeated question treatment (R) was identical to the O treatment except for the following: the subjects were informed that the experiment consists of a predetermined amount of rounds. *Rounds in our experiment implied that in case they decided not to destroy we would ask them up to 10 times if they wanted to change their decision*. The participants were not informed about the number of rounds, or times they were going to be asked, but only that there was a predetermined amount. If a participant chose to destroy, the game ended. There were up to ten rounds.

More specifically, at the start of the first round the subjects were asked if they wanted to reduce their partner's income at the cost of 1 Guilder. After all subjects made their choices, the experimenters record their choices in a separate sheet of paper. If they decided to reduce their partner's income, they did not make any further decision within the game, i.e., in the remainder of the 10 rounds. If they decided *not* to reduce their partner's income, in the following round they were asked if they were sure that they did not want to change their choice (i.e., reduce their partner's income). This question was posed to the subjects until they either changed their choice or until the experiment reached the final round. No feedback on their choices was provided to, or received about the actions of their partners, in between rounds.

Note that subjects could only destroy once throughout the game; once they chose to destroy, the game was effectively completed as far as they were concerned and they would have to wait.[15]

The aim of this treatment was to test for the effect of an indirect request or nudge. If this has any effect, it can be expected to be in the direction of increasing destruction. This is because it may start implicitly to build up on the pressure to destroy by repeating the question ten times.

Hypothesis 1. A higher destruction rate is expected in the R treatment than in the O treatment.

3.4. Repeated Compliance Constant pressure (RCC)

The Repeated Compliance Constant pressure treatment (RCC) extends the R treatment. At the end of the instructions the following sentence was added "it would be especially useful [to the experimenter] if you were to reduce your partner's income if you have not done so already. You are entirely free not to reduce if you wish".[16] This is our first direct pressure treatment where we explicitly request participants to destroy their partner's income.

This was the first subtle yet explicit cue from the experimenter to the subject to destroy. It was deliberately a subtle cue for three reasons. First, in many real world settings commands are phrased in similar equivalent subtle language, and we wanted our experiment to be applicable to more than contexts where a direct command is given.

Unlike Milgram (1963, 1974) or Cadsby et al. (2006), we did not provide an explicit reason to destroy in this treatment, though our instructions focused subjects on destroying as being particularly useful towards the experimenter without providing a justification.

We expect subjects to destroy more in the experiment as a result of the experimental authority cue being provided, and so in this treatment more than in the R treatment. This is because the cue makes both clear and stronger the demand from the authority to destroy.

Hypothesis 2. A higher destruction rate is expected in the RCC treatment than in the R treatment.

[15] We asked subjects whether they would have been willing to reduce their partners' income for an additional 5 Guilders at the cost of 1 Guilder, but this was a purely hypothetical and unincentivized question. See the online appendix for a brief analysis of the hypothetical responses.

[16] The wording in the square brackets was not included in the instructions, as it seemed obvious to us at the time that the benefit could only be towards the experimenter. Participants were fully informed of all stages of the experiment at the start and it was clear that their sole decision was to either destroy half of their partner income at an own monetary cost or not. Consequently, it was clear that there was no way it could benefit the participant. We chose precisely not to explain in what terms that would be useful to the experimenter as in a follow-up treatment we explicitly provide a justification for that request. This allows us later on to compare how providing a justification for compliance affects behavior. In terms of procedures, after the instructions were read aloud participants were given the option to ask clarification questions. In the rare case a participant would ask to whom it would be useful, it was clarified that it could not be useful to them or to any of the other participants. Lastly, as part of the end of experiment questionnaires, we added an open ended question, in which we asked the subjects what they believed was the scientific objective of the experiment. This question was provided in all the following treatments as well. As discussed later on no participant pointed that did not understand to whom usefulness were referred to. Nor did we have any complaints from participants that they found the phrasing misleading.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

3.5. Repeated Order (R-Order)

The Repeated Order treatment extends the R treatment to evaluate to what extend an explicit order, rather than a polite nudge as in RCC, affects compliance rates. In this treatment the participants were told: "We as experimenters give you the order now to reduce your partners' income. You now decide whether to obey to this order." This should increase compliance relative to the R treatment, but the effect relative to the RCC treatment is less clear. On the one side, a direct order can be expected to be more effective in obtaining obedience from subjects; on the other hand, the direct and less polite nature of the request may make it sound less reasonable, making compliance lower.[17]

Hypothesis 3. A higher destruction rate is expected in the R-Order treatment than in the R treatment.

3.6. Repeated Compliance (RC)

The Repeated Compliance (RC) treatment extends the RCC treatment. In this treatment, rounds 1, 4, 7 and 10 were marked with decision sheets provided on yellow paper. In this treatment the participant were told: "in the rounds with yellow instructions it would be especially useful if you were to reduce your partner's income if you have not done so already. You are entirely free not to reduce if you wish."

We expect that the 'yellow rounds' may be particularly effective in raising destruction. This is because, first, it is consistent with research on adaptive utility (e.g., Robson, 2001, for a discussion) showing how subjects are less likely to react with constant stimuli than they are with only occasional stimuli. Constant stimuli lead to adaptation and therefore ignoring, while intermittent signals are more likely to elicit a response.[18] Second, the 'yellow rounds' make more transparent and stronger what the experimenter demands in the rounds with peak pressure, while still allowing as many as four opportunities to destroy given our experimental setup.

Hypothesis 4. A higher destruction rate is expected in the RC treatment than in the RCC treatment.

Hypothesis 5. A higher destruction rate is expected in the RC treatment than in the R treatment.

3.7. Repeated Compliance Justified (RCJ)

The Repeated Compliance Justified treatment (RCJ) differs from the RC treatment in that, in addition to the same cue as in the RC treatment, participants are told that reducing their partner's income "would help us achieve a scientific objective of the experiment". Providing an explicit reason for destruction should increase destruction by increasing the moral legitimacy or providing an excuse to be nasty, either way therefore increasing further the compliance of participants to the authority (and/or increasing the likelihood of anticipated reciprocity).

Hypothesis 6. A higher destruction rate is expected in the RCJ treatment than in the RC treatment.

We assume that increasing pressure across treatments would lead to a cumulative effect on participants' compliance. Hence, given that we expect higher destruction in RC than RCC and a higher destruction in RCC than R, we also expect destruction to be higher in RCJ than in the R treatment.

Hypothesis 7. A higher destruction rate is expected in the RCJ treatment than in the R treatment.

3.8. Repeated Compliance 1-sided (RC-1)

The Repeated Compliance 1-sided treatment (RC-1) differs from the RC treatment in that half of the subjects were active and half were passive (though this terminology was not used in the instructions). Each active subject was matched with a passive partner but only the active subject made destruction decisions within the game, and this was common knowledge. Specifically, active subjects were told that "your partner answers some hypothetical questions but makes no decisions affecting your or his or her earnings".

The RC-1 treatment tests whether the cue by the authority induces destruction because of anticipated reciprocity. If destruction is entirely driven by an unconditional desire to comply there is no reason to expect that the destruction rates in RC-1 will differ from the destruction rates in RC. If however destruction is motivated by some anticipated reciprocity from partner's behavior, then the fact the partner can no longer affect the participant's earnings should lead to lower compliance and consequently destruction rates. We expect that at least some participants would be motivated by anticipated reciprocity and/or from a desire to ensure they will not receive a lower payoff than their partner leading to lower destruction rates on average in the RC-1 treatment than in the RC.

Hypothesis 8. A higher destruction rate is expected in the RC treatment than in the RC-1 treatment.

[17] See Pelligra et al. (2015) for an experiment showing how the perceived reasonableness of requests made by the experimenter on how much the trustee should return matters in the context of return rates in the context of trust games.

[18] As an example, this can be seen as a reason why alarms often have intermittent rather than constant beeping.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

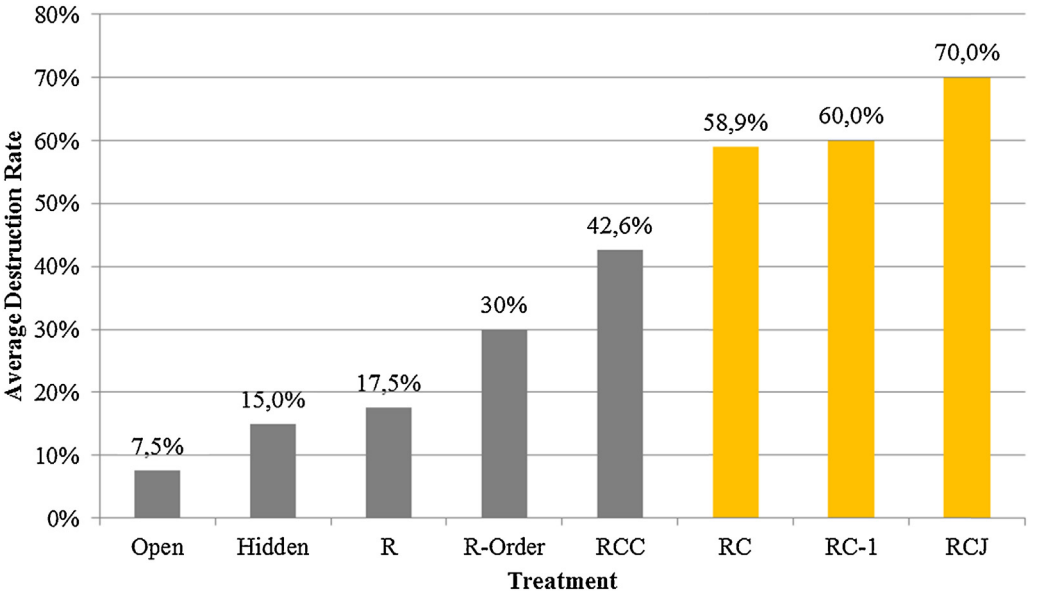

Fig. 2. Destruction rates across treatments. *Notes*. Yellow stands for treatments with 'yellow rounds' RC, RCJ and RC-1. Non-Yellow stands for treatments R and RCC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Regressions on destruction rates.

Variables	Regression 1		Regression 2	
	dy/dx	*p*	dy/dx	*p*
O	−0.161 (0.119)	0.177	−0.190 (0.048)	0.111
H	−0.032 (0.107)	0.762	−0.048 (0.053)	0.652
RCC	0.238*** (0.090)	0.008	0.233*** (0.084)	0.009
R-Order	0.131 (0.099)	0.187	0.090 (0.075)	0.367
RC	0.370**** (0.085)	0.000	0.364**** (0.083)	0.000
RCJ	0.465**** (0.090)	0.000	0.440**** (0.083)	0.000
RC-1	0.378**** (0.092)	0.000	0.370**** (0.075)	0.000
British			−0.097** (0.064)	0.044
Gender			0.051 (0.049)	0.281
Age			−0.003 (0.006)	0.566
Economics			−0.071 (0.067)	0.276
Atheist			−0.054 (0.088)	0.249
N	350		350	
Pseudo R^2	0.159		0.176	
Prob > X^2	0.000		0.000	

Notes: Marginal effects of Probit regressions with robust standard errors.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.
**** $p < 0.001$.
Other than treatment variables, regressions include age (subtracted from mean age), gender (= 1 for women), economics background (= 1 if applicable), nationality (British = 1 for British subjects, Atheist = 1 for atheist and agnostic subjects).

As no anticipatory retaliation was possible in this treatment, there was no question regarding expected destruction from the partner at the end of the experiment.

4. Results

4.1. Testing the hypotheses

Fig. 2 presents the average destruction rate across treatments. Table 1 presents the results of Probit regressions on whether subjects choose to destroy (= 1) or not (= 0) and their implications in terms of overall winning probabilities for each treatment. The regressions employ dummy variables for the experimental treatments; the R treatment was used as baseline. We included dummies for British nationality, gender, age economics students and atheist as controls. Table 2 reports Wald

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

Table 2
Wald tests on coefficients of regressions on destruction rates.

Wald tests (p-values)		
	Regression 1	Regression 2
Open = Hidden	0.291	0.239
R-Order = RCC	0.212	0.097
RC = RCC	0.087	0.089
RC = R-Order	0.005	0.002
RC = RC-1	0.916	0.937
RC = RCJ	0.266	0.370

tests of equality of coefficients for our treatment variables. The result supports the interpretation of destruction as compliance to the authority, whereas the demographic variables are generally insignificant. We now consider the hypotheses and the key evidence.

Result 1. Against Hypothesis 1, there is no statistically significant difference between destruction rates in the H, O and the R treatments.

Evidence. Destruction in R treatment is roughly the same as those in the H treatment and both are qualitatively above O, which is what we would expect and descriptively in line with the results of Abbink and Herrmann (2011). The regression coefficient on O is not statistically significant for either regression 1 ($p = 0.18$) or 2 ($p = 0.11$). The same results can be obtained in a simple bivariate test by comparing destruction rate proportions in the Open and Hidden treatments using a Fisher's exact test ($p = 0.24$) or between the Open and R ($p = 0.31$) treatments.[19] Overall, repeating the question again and again does not make a significant difference, with the destruction rate remaining below 20%.

Result 2. In support of Hypothesis 2, the destruction rate is statistically significantly higher in the RCC treatment than in the R treatment. The difference is large.

Evidence. Fig. 2 shows how the RCC destruction rates are more than double than that in the R treatment. A bivariate Fisher's exact test is significant ($p = 0.01$). The same is true for our regression coefficients which are positive and highly statistically significant ($p < 0.01$).

Result 3. In contrast to Hypothesis 3, the destruction rate is not statistically significant different between the R and the R-Order treatment.

Evidence. Fig. 2 shows that the destruction rate in the R-Order is between the R and RCC treatments. A bivariate Fisher's exact test between the R and the R-Order values is not statistically significant ($p = 0.29$). The regression coefficients are not statistically significant in neither regression 1 ($p = 0.19$) nor in regression 2 ($p = 0.37$).

That said, the R-Order destruction rate is also not statistically significant from that in the RCC treatment using a Fisher's exact test ($p = 0.28$) or in regression 1 (Wald test, $p = 0.21$), while being marginally significant in regression 2 ($p = 0.097$).

Result 4. In support of Hypothesis 4, the destruction rate is weakly statistically significantly higher in the RC treatment than in the RCC treatment.

Evidence. As shown by Fig. 2, quantitatively, the destruction rate in RC is about 16% more than in RCC, suggesting that providing peak pressure at intervals makes a difference. A bivariate Fisher's exact test yields a weak statistical significance in a one-sided test ($p = 0.06$). Conducting Wald tests between RCC and RC suggest a weak statistical significant difference in both regression 1 ($p = 0.087$) and regression 2 ($p = 0.089$).[20] The supplementary analysis of Section 4.2 provides complementary evidence.

Result 5. In support of Hypothesis 5, the destruction rate is statistically significantly higher in the RC treatment than in the R treatment. The difference is large.

Evidence. Fig. 2 shows how the destruction rate more than triples when moving from the R treatment (17.5%) to the RC treatment (58.9%), suggesting a cumulating effect between cue and peak pressure. This is significant in a bivariate Fisher's exact test ($p < 0.001$) and in the regression analysis. The estimated coefficients on RC are positive and statistically significant in both regression 1 ($p < 0.001$) and regression 2 ($p < 0.001$).

Result 6. Against Hypothesis 6, there is no statistically significant difference between destruction rates in the RC and the RCJ treatments.

Evidence. While the destruction rate is as high as 70% in the RCJ treatment, the difference from the RC treatment is not enough to achieve statistical significance either in a bivariate Fisher's exact test ($p = 0.29$) or in the Wald tests ($p = 0.216$ in regression 1 and $p = 0.37$ in regression 2). Providing an explicit reason for destruction does not seem to make a difference.

Result 7. In support of Hypothesis 7, the destruction rate is statistically significantly higher in the RCJ treatment than in the R treatment. The difference is large.

[19] All p-values reported in this paper are two tailed unless specified otherwise.
[20] Note that the evidence is stronger when comparing RCC with R-Order. For example, Wald tests between the relevant coefficients result in $p = 0.005$ for regression 1 and $p = 0.002$ for regression 2.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

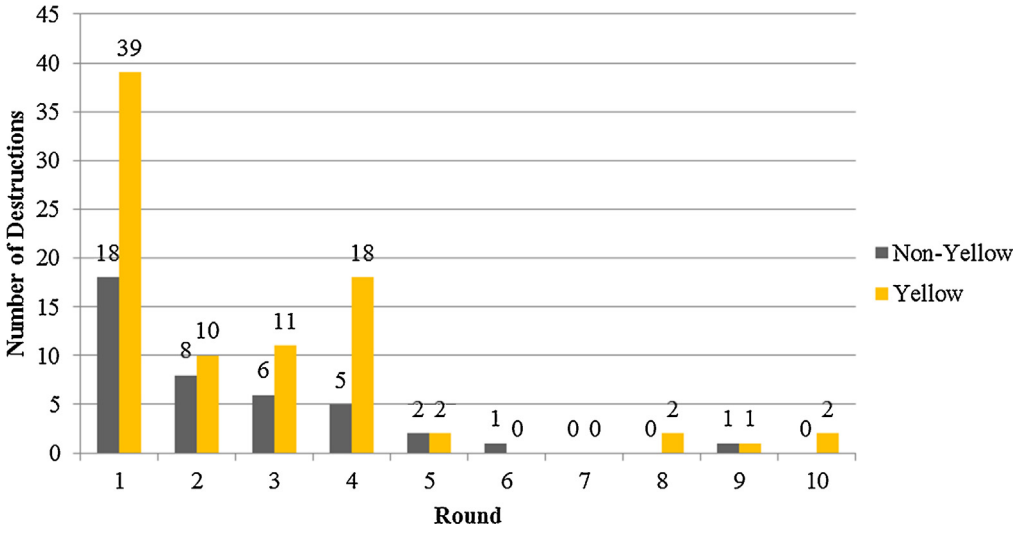

Fig. 3. Destruction rates per round. *Notes*. Yellow stands for treatments with 'yellow rounds' RC, RCJ and RC-1. Non-Yellow stands for treatments R, RCC and R-Order. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Destruction rates in each round.

Round	R	R-Order	RCC	RC	RCJ	RC-1
1	5.00%	17.5%	22.50%	25.00%	37.50%	25.00%
2	2.63%	6.06%	16.13%	9.52%	16.00%	6.67%
3	5.40%	3.22%	11.53%	10.52%	14.28%	14.29%
4	0.00%	3.22%	17.39%	20.58%	27.78%	25.00%

Notes: The destruction rate in round 1 is the proportion of all subjects who have destroyed in round 1. The destruction rates for rounds 2, 3 and 4 are conditional on the destruction in the previous rounds: specifically, they are the proportions of all subjects who have destroyed in the given round conditional on them not having destroyed in the previous round(s).

Evidence. The destruction rate is four times as high in the RCJ treatment as in the R treatment. It should then come to no surprise that this difference is statistically significant in a bivariate Fisher's exact test ($p < 0.001$) or in the regression analysis, where the estimated coefficients for the RCJ dummy yield $p < 0.001$ for both regressions 1 and 2.

Result 8. Against Hypothesis 8, there is no statistically significant difference between destruction rates in the RC and the RC-1 treatments.

Evidence. Fig. 2 already provides the answer by showing virtually identical destruction rates (60% in RC-1 vs 58.9% in RC), not statistically significantly different in a bivariate Fisher's exact test ($p = 1.00$). The Wald tests between the RC and RC-1 dummy in regressions 1 ($p = 0.916$) and 2 ($p = 0.937$) further confirm this. Overall, this points to compliance to authority rather than anticipated reciprocity by partners as the driver of destruction.[21]

4.2. Supplementary analysis

Time of destruction. It is interesting to see when the decision to destroy takes place over the ten rounds of most treatments (i.e., all treatments other than O and H). Fig. 3 presents destruction rates per round for the 'yellow rounds' repeated play treatments where peak pressure is applied to destroy at given points in time (RC, RCJ, and RC-1) and for treatments where constant pressure was applied (R, RCC and R-Order). Note that subjects can only destroy once, and therefore those who have already destroyed in the early rounds cannot destroy further in the following rounds; effectively, later decisions are conditional on not having destroyed before, reflecting a sample selection increasingly composed of subjects who are less willing to comply with the experimental cue provided.

Fig. 3 shows that almost all of the destruction takes places by the 4th round, and Table 3 provides a breakdown of destruction rates (conditional on the previous round's destruction from round 2 onwards) up to round 4. Despite the declining

[21] It is interesting to compare and contrast our simple between treatments test on anticipation of reciprocity with the use of belief elicitation data, which we also collected at the end of the experiment for all treatments other than RC-1 (where it obviously could not be formulated). There tends to be a positive correlation between destruction and stated belief about the destruction of the other partner, though there is no clear pattern to it across treatments and it may be largely driven by noise (Spearman $\rho = 0.854$, −0.009, 0.223, −0.126, 0.297, 0.480, 0.180, respectively with $p < 0.001$, $p = 0.96$, 0.17, 0.03, 0.44, <0.01, and = 0.195 in treatments O, H, R, R-Order, RC, RCJ and RCC respectively). It is possible that having decided to destroy might make it more likely for subjects to believe the partner will, e.g., out of self-image and cognitive dissonance concerns.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

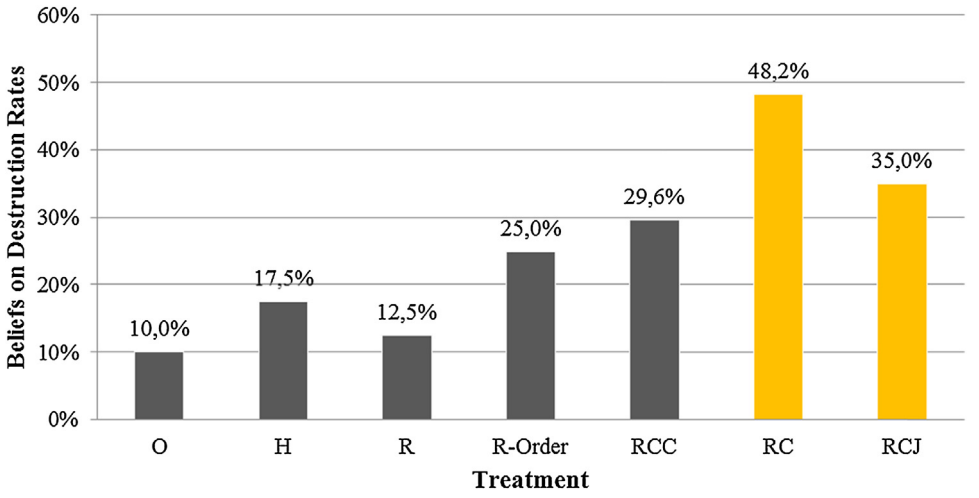

Fig. 4. Beliefs on destruction rates accross treatments. *Notes*. Yellow stands for treatments with 'yellow rounds' RC, RCJ and RC-1. Non-Yellow stands for treatments R and RCC. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Regressions on destruction rates with beliefs.

Variables	Regression 3		Regression 4	
	dy/dx	p	dy/dx	p
Open	−0.168 (0.105)	0.107	−0.187* (0.105)	0.074
Hidden	−0.038 (0.102)	0.707	−0.051 (0.062)	0.612
RCC	0.189** (0.085)	0.026	0.187** (0.079)	0.026
R-Order	0.099 (0.095)	0.296	0.075 (0.075)	0.432
RC	0.276 (0.082)	0.001	0.277*** (0.072)	0.001
RCJ	0.403**** (0.083)	0.000	0.391**** (0.050)	0.000
Beliefs			0.196**** (0.008)	0.000
British			−0.066 (0.064)	0.175
Gender			0.023 (0.051)	0.631
Age			−0.003 (0.006)	0.619
Economics			−0.061 (0.070)	0.380
Atheist			−0.030 (0.096)	0.536
N	310		310	
Pseudo R^2	0.203		0.211	
Prob > X^2	0.000		0.000	

Notes: Marginal effects of Probit regressions with robust standard errors.
* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.
**** $p < 0.001$.
Other than treatment variables, regressions include age (subtracted from mean age), gender (= 1 for women), economics background (= 1 if applicable), nationality (British = 1 for British subjects, Atheist = 1 for atheist and agnostic subjects).

trend across rounds due to early decisions to destroy and sample selection, in the 'yellow rounds' treatments with peaks in pressure to destroy there is a spike in destruction rates in round 4. This is in line with our expectations, and complements Result 4 above, as round 4 was one of the rounds in which we asked the subjects explicitly to reduce their partners' income.

In treatments other than O or H, as subjects know that there are multiple rounds, subjects can defer destruction to after round 1, and many do so.[22]

Data on beliefs. Fig. 4 presents the beliefs on average destruction rates by partners across treatments. Table 4 presents the results of Probit regressions on whether subjects choose to destroy or not including a dummy for beliefs instead of RC-1.

The first observation that can be drawn from Fig. 4 is that on average the subjects expected a lower destruction from their partners than what was actually observed (Wilcoxon $p = 0.02$). Nevertheless, beliefs predict the decision to destroy in both regressions reported in Table 4 ($p < 0.001$). With respect to changes across treatments, beliefs on destruction between

[22] We ran Probit regression analysis specifically on round 1 destruction (see online appendix). The coefficients on the O and R dummies are negative and statistically significant relative to the RC baseline; that on H is not.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

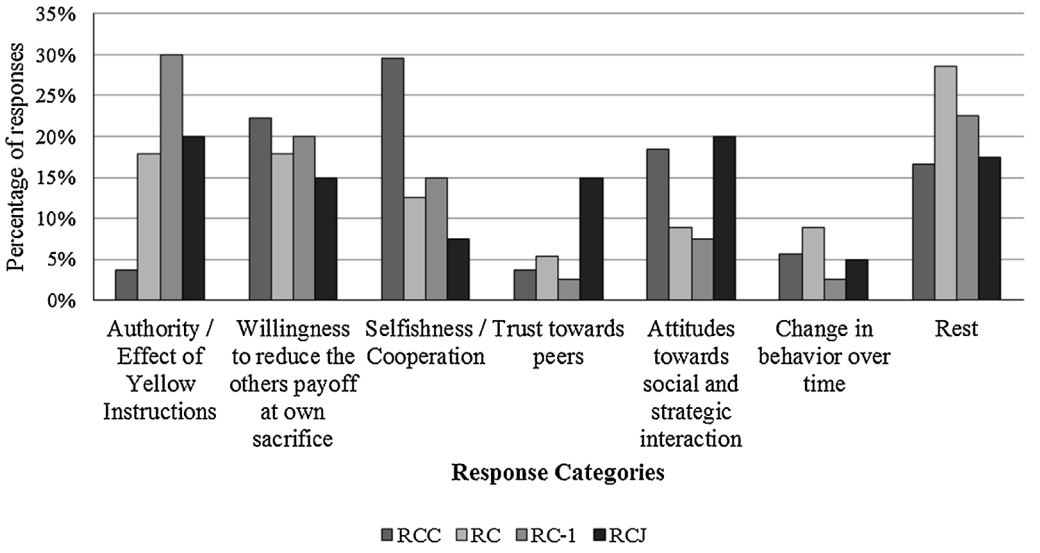

Fig. 5. Proportion of qualitative responses per category for each treatment. *Notes*. The responses were to a question on what subjects thought was the objective of the experiment.

O, H and R are similar to each other.[23] However, this is not true for R and RCC or RCC and RCJ. In these cases, in line with the increase in destructions rates there is an increase on beliefs on partner's destruction.[24]

Qualitative data. As noted earlier, in the treatments with an explicit cue by the authority/experimenter to destroy (RC, RCC, RC-1 and RCJ),[25] at the end of the experiment we asked subjects what they thought the objective of the experiment was.[26] The answers were then grouped in categories by research assistants themselves not informed of the objectives of the experiment.[27] The answers were divided into 7 broad categories: authority/effect of yellow instructions, willingness to reduce the others payoff at an own sacrifice, selfishness vs cooperation, trust towards peers, attitudes towards social and strategic interaction, and change in behavior over time, and rest.[28] Fig. 5 displays the results of the classification.

Unsurprisingly, while between 15% and 30% of the subjects in the treatments with 'yellow rounds' peaks of pressure picked 'authority/effect of yellow instructions' as the objective of the experiment, less than 5% did so in the RCC treatment with constant pressure and no 'yellow rounds.'[29] The modal answer in RCC switched to 'selfishness/cooperation', again a plausible response insofar as subjects can avoid destroying each other.[30]

There seem to be three useful messages from the qualitative data. First, all of the six substantive categories pick up on potentially relevant aspects of the decision problem, displaying no evidence of confusion about the decision problem. Second, we had no evidence, either from these categories or looking at the residual category, of subjects wanting to destroy money in order to indirectly return money to the experimenter, which would be a kind of house money effect.[31] Third,

[23] There is no statistically significant difference between O and H or O and R (Fisher's exact test, $p = 0.51$ and $p = 1$ respectively).
[24] A weakly statistically significant difference is observed between R and RCC and RCC and RCJ ($p = 0.078$ and $p = 0.053$ respectively). The difference between RC and RCJ is not statistically significant (Fisher's exact test, $p = 0.216$).
[25] The R-Order treatment was conducted later after a suggestion by a referee and we did not collect qualitative data for this treatment.
[26] Such data obviously has limitations, as discussed in Zizzo (2010).
[27] The answers were first grouped into categories by a research assistant with no prior knowledge of the aims of the experiment or experimental design. Afterwards a second research assistant, with no prior knowledge of the aims of the experiment, received the previously created seven categories along with the subjects' responses and was asked to independently match every answer with one of the categories. Finally, the assistants have met up to reconcile any discrepancy. Our procedure follows a tradition of coding in qualitative research (e.g., Saldaña, 2012), adapted to eliminate any risk of any research project related in the coding of categories; see Bourgheas et al. (2013) for another example within experimental economics.
[28] *Authority/effect of yellow instructions* referred to subjects who believed that the aim of the experiment was either the influence of authority upon them or the influence of the 'yellow paper' rounds. *Willingness to reduce the others payoff at own sacrifice* saw the experiment being about subjects reducing their partner's income despite there was no monetary incentive for them to do so. *Selfishness/cooperation* referred to whether people are selfish or willing to cooperate. Similarly, *Trust* was about subjects being able to trust each other not to destroy. *Attitudes towards social and strategic interaction* was about when subjects argued that the aim of the experiment relates to strategic behavior, either on what the other subject will do or what he or she would do if he or she was in their place (RC-1 treatment). *Change in behavior over time* answers were about the experiment being about change in choice if they are asked repetitively if they will change their choice. *Rest* was the residual category and included non-responders.
[29] The difference between the RCC and the remaining treatments is statistically significant (Fisher exact tests, $p = 0.017$, 0.014 and <0.001 for RC, RCJ and RC-1 respectively).
[30] This difference is also statistically significant across all treatments (Fisher's exact tests, $p = 0.024$, 0.007 and 0.07 for RC, RCJ and RC-1 respectively).
[31] In support of this negative finding, in a different experiment by the second author with Piers Fleming, we offer the opportunity to subjects to materially destroy University cafes coupons by putting them through a shredder, which rules out this house money effect explanation; however, many do when provided with a cue in the form of what subjects did in a previous treatment (Fleming and Zizzo, 2015).

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

there is no evidence that subjects wanted to be altruistic towards the experimenter as such; the focus even of the answers classified as related to 'selfishness/cooperation' was with respect to partners.

5. Discussion and conclusions

We found that a norm of compliance exists, independently of any social norm to be prosocial. A limited nudge, without an explicit justification, can induce compliance from 60% of the subjects.[32] Our 60% destruction rate compares with previous experiments on antisocial behavior that have achieved destruction rates between 10% and 40%.[33] While unequivocal, our cue was not phrased directly; indirect cues of this kind are arguably pervasive in both private and public organizations.[34] Phrasing the cue as a direct order was instead not particularly helpful. We conjecture that, in this case, the impoliteness and evident arbitrariness of the requests compensated for any additional effectiveness deriving from the directness of the cue.[35] Clearly, more research on different formulations, and the norms that these different formulations appeal to, would be an interesting area for further research and relevant for public policy. Additionally, in a taxation context, it may be relevant to identify the most effective nudges (for example, in the form of advertising) to elicit tax compliance.

In the real world, employees may be compliant because there are clear monetary benefits from doing so (see Robin et al., 2012). However, the fact that participants had no monetary incentives to be compliant (but rather the opposite) in this experiment, is consistent with the existence of a norm of compliance. The verb 'useful' employed in most treatments may have suggested to subjects that there was some value for the experimenter from destroying. It does not explain however why the subjects should do something of value for the experimenter when they are equally told that they are free to do whatever they like.[36] This is where the adherence to a norm to comply towards the (trusted) authority comes in.

Our results cannot be explained by purely repeating the question again and again, since by doing so destruction rates remained below 20%. They also cannot be explained by reciprocal expectations of destruction from the partner, since, even in the absence of the possibility of reciprocity, we observe around a 60% destruction rate. Providing pressure at peak intervals does help the authority to induce more compliance. In a tax context, one could think of adverts encouraging tax compliance say in the month before the official deadline.

Arrow (1974) argued that compliance to authority, whether this is the authority of law or to a superior in an organization, is due to expectations that others will also comply. This is an important idea both for public policy and for the management of organizations. Our data on beliefs provides some support to this hypothesis. On average participants expected significantly fewer people to comply than the amount observed, but we do observe that the greater the pressure to comply more participants expected others to comply and destruction was higher. Our qualitative data suggests that subjects generally did understand the nature of the task and that their compliance was not driven by a desire to return money to the experimenter. Of course, qualitative data has limitations and should be taken with caution.

A deliberate and explicit experimenter demand (Zizzo, 2010) was used as a tool to study the effect of authority. Providing the role of an authority to a subject would have been inappropriate as it would have mixed up an effect of authority with that of peer pressure, preventing the identification of the former in isolation. Our choice of the experimenter as the source of authority is of course the same made by Milgram (1963, 1974). Relative to Milgram (1963, 1974), our study checked how far one can go with authority when considerably weaker and less explicit cues to obey are employed, and such weaker and less explicit cues also more in line with the reality of the functioning of authority in the every-day life of organizations.

Overall, our experiment suggests that a norm of compliance towards the authority should not be neglected in principal-agent modeling, nor should it be neglected as a powerful managerial and policy tool. It is arguable that, even in the lack of economic incentives, individuals may tend at least to some degree to comply with requests, though a more direct request is not necessarily more effective than a less direct request. This is exploited by economic organizations big and small as a management tool. For example, it may help strategic delegation which is advantageous to principals in handling conflict and contests (Wärneryd, 2012). As another example, it may be one reason (complementary to others) why marginal wage

[32] It is interesting to note that, in the different context of prosocial behavior ultimately in one's own self-interest, Silverman et al. (2014) also find an effect for a cue but not an additional effect for a justification, at least when not *also* combined with the threat of a financial penalty, and equally Sonntag and Zizzo (2015) do not find a significant aggregate effect of a justification.

[33] This has been achieved using methods such as the ability to hide due to random destruction (Abbink and Herrmann, 2011; Abbink and Sadrieh, 2009) or the introduction of pointless prizes (Abbink and Herrmann, 2009).

[34] An indirect cue is also useful within the experimental context to address the potential criticism that subjects may be destroying simply because they *have* to follow the instructions on what to do in the experiment. This is particularly true as we made simultaneously clear that they were free to do whatever they wanted.

[35] A reviewer suggested an alternative interpretation: namely, that the 30% destruction rate in the R-Order treatment is our best estimate of a pure obedience effect. We feel this may be inaccurate in both directions. It may be an over-estimate because destruction effects in the R treatment may be due to pure antisocial preferences (as per Abbink and Sadrieh, 2009), and, if so, only the difference in destruction rates between the R and R-Order treatments (12.5%) could be purely identified as an obedience effect. It may also be an under-estimate because the RCC treatment may also be picking up a pure obedience effect.

[36] One possible response would be that the experimenter is not an anonymous person and subjects may have a repeated relationship with the lab. This would be a problem if subjects felt that they had to be afraid of any repercussions if they did not follow our cue. However, as far as we can tell not a single qualitative comment by any subject reflected a fear of this kind. In addition, if such fear were true, it should have been heightened in the R-Order treatment, leading to more destruction relative to RCC, which is in contrast to our results. The UEA laboratory is well established, with many experiments run each year, an established ethical protocol and a clear reputation that what happens in an experiment remains in the experiment.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

incentives are found less in organizations than they should under principal-agent modeling with self-interest (Milgrom and Roberts, 1992). Where authority is in the wrong hands, cues to engage in aggressive or harassing behavior do not have to go as far as giving explicit orders (the additional effectiveness of which is anyway unclear based on our results). There seems to be evidence of this in organizations (Ashford and Anand, 2003; Brief et al., 1995; Darley, 2001); a desire to influence, as recently found experimentally by Sadrieh and Schröder (2012), may be the counterpart of the inclination to comply.[37]

Within a tax context, it is useful to identify how effective a norm of compliance can be when isolated from prosocial norms that may also support tax compliance. As noted earlier, this may inform policies, such as advertising campaigns, that try to leverage on such a norm of compliance. The converse of our finding, of course, is that some 30–40% of subjects were resistant to the cues by the authority, even when phrased reasonably, repeated ten times, under intervals of peak pressure and with an explicit justification. If our sample is more broadly representative (which, of course, is a significant if), this suggests that a different type of intervention is likely to be needed to reach out to about 1/3 of taxpayers.

Acknowledgments

Financial support from the ESRC (NIBS Grant ES/K002201/1) and the University of East Anglia is gratefully acknowledged. We thank Abigail Barr, Jason Dana, Enrique Fatas, Werner Güth, David Masclet, Karim Sadrieh, Lilia Zhurakhovska, an editor, two reviewers, and participants to presentations in Bath, Birmingham, Brussels, Cologne, Coventry, Lyon, Maastricht, Norwich, Nuremberg, Rennes and Tilburg for useful advice, and Axel Sonntag and Ailko van der Veen for research assistance. The usual disclaimer applies. An online appendix, inclusive of experimental instructions, is available. The experimental dataset has been offered to the UK Data Archive.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2015.09.016.

References

Abbink, K., Herrmann, B., 2009. Pointless Vendettas. University of East Anglia CBESS, Working Paper 2009-02, Norwich.
Abbink, K., Herrmann, B., 2011. The moral costs of nastiness. Econ. Inq. 49, 631–633.
Abbink, K., Sadrieh, A., 2009. The pleasure of being nasty. Econ. Lett. 105, 306–308.
Abbink, K., Brandts, J., Herrmann, B., Orzen, H., 2010. Inter-group conflict and intra-group punishment in an experimental contest game. Am. Econ. Rev. 100, 420–447.
Aghion, P., Tirole, J., 1997. Formal and real authority in organizations. J. Polit. Econ. 105, 1–29.
Alm, J., Torgler, B., 2006. Culture differences and tax morale in the United States and in Europe. J. Econ. Psychol. 27 (2), 224–246.
Anderhub, V., Gächter, S., Königstein, M., 2002. Efficient contracting and fair play in a simple principal-agent experiment. Exp. Econ. 27, 5–27.
Andreoni, J., Bernheim, D.B., 2009. Social image and the 50–50 norm: a theoretical and experimental analysis of audience effects. Econometrica 77, 1607–1636.
Ariely, D., Bracha, A., Meier, S., 2009. Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am. Econ. Rev. 99, 544–555.
Arrow, K.J., 1974. The Limits of Organization. Norton, New York.
Asch, S., 1955. Opinions and social pressure. Sci. Am. 193, 31–35.
Ashford, B.E., Anand, V., 2003. The normalization of corruption in organizations. In: Staw, B.M., Kramer, R.M. (Eds.), Research in Organizational Behavior. JAI, Greenwich, CT.
Bardsley, N., Sausgruber, R., 2005. Conformity and reciprocity in public good provision. J. Econ. Psychol. 26, 664–681.
Benabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96, 1652–1678.
Bicchieri, C., 2006. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, Cambridge.
Bicchieri, C., Xiao, E., 2009. Do the right thing: but only if others do so. J. Behav. Decis. Mak. 22, 191–208.
Blass, T., 1999. The Milgram paradigm after 35 years: some things we now know about obedience to authority. J. Appl. Soc. Psychol. 29, 955–978.
Bone, J., Drouvelis, M., Ray, I., 2012. Following Recommendations to Avoid Coordination-failure in 2 × 2 Games. University of Birmingham Economics, Discussion Paper 12-04.
Bourgheas, S., Nieboer, J., Sefton, M., 2013. Risk-Taking in Social Settings: Group and Peer Effects. University of Nottingham CEDEX, Discussion Paper 2013-01.
Brandts, J., Cooper, D., Fatas, E., 2007. Leadership and overcoming coordination failure with asymmetric costs. Exp. Econ. 10, 269–284.
Brief, A.P., Buttram, R.T., Elliot, J.D., Reizenstein, R.M., McCline, R.L., 1995. Releasing the beast: a study of compliance with orders to use race as a selection criterion. J. Soc. Issues 51, 177–194.
Cadsby, C.B., Maynes, E., Trivedi, V.U., 2006. Tax compliance and obedience to authority at home and in the lab: a new experimental approach. Exp. Econ. 9, 343–359.
Case, A.C., Katz, L.F., 1991. The Company You Keep: The Effects of Family and Neighbourhood on Disadvantaged Youths. NBER, Working Paper 3705.
Cason, T.N., Mui, V.-L., 1998. Social influence in the sequential dictator game. J. Math. Psychol. 42, 248–265.
Cason, T.N., Sharma, T., 2007. Recommended play and correlated equilibria: an experimental study. Econ. Theory 33, 11–27.
Cialdini, R.B., Goldstein, N.J., 2004. Social influence: compliance and conformity. Annu. Rev. Psychol. 55, 591–621.
Croson, R., Marks, M., 2001. The effect of recommended contributions in the voluntary provision of public goods. Econ. Inq. 39, 238–249.
Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ. Theory 33, 67–80.

[37] One specific further area for future research could revolve around understanding the interaction between social norms and obedience to authority. Another one could be the extent to which compliance utility has been internalized by subjects as part of their self-image or is reliant on a simple rule of thumb. In both of these cases, compliance utility would not depend on knowing that the authority may learn about their actions.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016

Darley, J.M., 2001. The dynamics of authority in organization. In: Darley, J.M., Messick, D.M., Tyler, T.R. (Eds.), Social Influences on Ethical Behavior in Organizations. Erlbaum, Mahwah, NJ/London, pp. 37–52.
Denant-Boemont, L., Masclet, D., Noussair, C.N., 2007. Punishment counterpunishment and sanction enforcement in a social dilemma experiment. Econ. Theory 33, 145–167.
Duffy, J., Feltovich, N., 2008. Correlated Equilibria, Good or Bad: An Experimental Study. University of Pittsburgh and University of Aberdeen, Working Paper.
Engel, C., Zhurakhovska, L., 2012. Small Harm on an Innocent Outsider as a Lubricant of Cooperation – An Experiment. MPI Collective Goods, Preprint No. 2012/2.
Fehr, E., Klein, A., Schmidt, K., 2007. Fairness and contract design. Econometrica 75, 121–154.
Feld, L.P., Frey, B.S., 2002. Trust breeds trust: how taxpayers are treated. Econ. Gov. 3, 87–99.
Fleming, P., Zizzo, D.J., 2015. A simple stress test of experimenter demand effects. Theory Decis. 78, 219–231.
Frey, B.S., Meier, S., 2004. Social comparison and pro-social behavior: testing conditional cooperation in a field experiment. Am. Econ. Rev. 94, 1717–1722.
Glazer, A., Konrad, K.A., 1996. A signalling explanation for charity. Am. Econ. Rev. 86, 1019–1028.
Grossman, S.J., Hart, O.D., 1986. The costs and benefits of ownership: a theory of vertical and lateral integration. J. Polit. Econ., 691–719.
Güth, W., Königstein, M., Kovács, J., Zala-Mezo, E., 2001. Fairness within firms: the case of one principal and multiple agents. Schmalenbach Bus. Rev. 53, 82–101.
Güth, W., Levati, M.V., Sutter, M., van der Heijden, E., 2007. Leading by example with and without exclusion power in voluntary contribution experiments. J. Public Econ. 91, 1023–1042.
Hart, O., Moore, J., 1990. Property rights and the nature of the firm. J. Polit. Econ. 98, 1119–1158.
Herrmann, B., Orzen, H., 2008. The Appearance of Homo Rivalis: Social Preferences and the Nature of Rent Seeking. CeDEx, Discussion Paper 2008-10.
Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Jones, S.R.G., 1984. The Economics of Conformism. Basil Blackwell, Oxford.
Karakostas, A., Sonntag, A., Zizzo, D., 2013. Efficiency and Fairness in Revenue Sharing Contracts, Available at SSRN.
Kawaguchi, D., 2004. Peer effects on substance use among American teenagers. J. Popul. Econ. 17, 351–367.
Keizer, K., Lindenberg, S., Steg, L., 2008. The spreading of disorder. Science 322 (5908), 1681–1685.
Kilham, W., Mann, L., 1974. Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm. J. Pers. Soc. Psychol. 29, 696–702.
Krupka, E., Weber, R.A., 2009. The focusing and informational effects of norms on pro-social behavior. J. Econ. Psychol. 30, 307–320.
Landry, C., Lange, A., List, J.A., Price, M.K., Rupp, N.G., 2006. Toward an understanding of the economics of charity: evidence from a field experiment. Q. J. Econ. 121, 747–782.
Lopez-Perez, R., 2008. Aversion to norm-breaking: a model. Games Econ. Behav. 64, 237–267.
Lundborg, P., 2006. Having the wrong friends? Peer effects in adolescent substance use. J. Health Econ. 25, 214–233.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60, 531–542.
Meeus, W.H.J., Raaijmakers, Q.A.W., 1995. Obedience in modern society: the Utrecht studies. J. Soc. Issues 51, 155–175.
Milgram, S., 1963. Behavioral study of obedience. J. Abnorm. Soc. Psychol. 67, 371–378.
Milgram, S., 1974. Obedience to Authority: An Experimental View. Harper and Row, New York.
Milgrom, P., Roberts, J., 1992. Economics, Organization and Management. Prentice Hall, Englewood Cliffs, NJ.
Moxnes, E., van der Heijden, E., 2003. The effect of leadership in a public bad experiment. J. Confl. Resolut. 47, 773–795.
Nikiforakis, N., 2008. Punishment and counter-punishment in public good games: can we really govern ourselves? J. Public Econ. 92, 91–112.
Nikiforakis, N., Engelmann, D., 2008. Feuds in the Laboratory? A Social Dilemma Experiment. University of Melbourne, Department of Economics, Research Paper 1058.
Orne, M.T., 1962. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. Am. Psychol. 17, 776–783.
Pelligra, V., Reggiani, T., Zizzo, D.J., 2015. Responding to (Un)Reasonable Demands. Working Paper (in preparation).
Powell, L.M., Tauras, J.A., Ross, H., 2005. The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behaviour. J. Health Econ. 24, 950–968.
Reicher, S.D., Haslam, S.A., Smith, J.R., 2012. Working toward the experimenter: reconceptualizing obedience within the Milgram paradigm as identification-based followership. Perspect. Psychol. Sci. 7, 315–324.
Robin, S., Rusinowska, A., Villeval, M.-C., 2012. Ingratiation and Favoritism: Experimental Evidence. IZA, Discussion Paper No. 6530.
Robson, A.J., 2001. The biological basis of economic behavior. J. Econ. Lit. 39, 11–33.
Rosnow, R.L., Rosenthal, R., 1997. People Studying People: Artifacts and Ethics in Behavioral Research. Freeman, New York.
Sadrieh, A., Schröder, M., 2012. The Desire to Influence Others. University of Magdeburg, Discussion Paper.
Saldaňa, J., 2012. The Coding Manual for Qualitative Researchers. Sage, Los Angeles.
Shanab, M.E., Yahya, K.A., 1977. A behavioral study of obedience in children. J. Pers. Soc. Psychol. 35, 530–536.
Shanab, M.E., Yahya, K.A., 1978. A cross-cultural study of obedience. Bull. Psychon. Soc. 11, 267–269.
Shang, J., Croson, R., 2009. Field experiments in charitable contribution: the impact of social information on the voluntary provision of public goods. Econ. J. 119, 1422–1439.
Silverman, D., Slemrod, J., Uler, N., 2014. Distinguishing the role of authority in and authority to. J. Public Econ. 113, 32–42.
Simon, H.A., 1946. The proverbs of administration. Public Adm. Rev. 6 (1), 53–67.
Simon, H.A., 1979. Rational decision making in business organizations. Am. Econ. Rev., 493–513.
Sonntag, A., Zizzo, D.J., 2015. Institutional authority and collusion. Southern Econ. J. 82, 13–37.
Stöber, J., 2001. The social desirability scale-17 (Sds-17) – convergent validity discriminant validity, and relationship with age. Eur. J. Psychol. Assess. 17, 222–232.
Thaler, R.H., Sunstein, C.R., 2008. Nudge. Yale University Press.
Thöni, C., Gächter, S., 2012. Peer Effects and Social Preferences in Voluntary Cooperation. IZA, Discussion Paper 6277.
Van der Heijden, E., Potters, J., Sefton, M., 2008. Hierarchy and opportunism in teams. J. Econ. Behav. Org. 69, 39–50.
Wärneryd, K., 2012. The evolution of preferences for conflict. Econ. Lett. 116 (1), 102–104.
Zafar, B., 2011. An experimental investigation of why individuals conform. Eur. Econ. Rev. 55, 774–798.
Zizzo, D.J., 2003. Money burning and rank egalitarianism with random dictators. Econ. Lett. 81 (2), 263–266.
Zizzo, D.J., 2010. Experimenter demand effects in economic experiments. Exp. Econ. 13, 75–98.
Zizzo, D.J., Fleming, P., 2011. Can experimental measures of sensitivity to social pressure predict public good contribution? Econ. Lett. 111, 239–242.
Zizzo, D.J., Oswald, A.J., 2001. Are people willing to pay to reduce others' incomes? Ann. Econ. Stat., 39–65.

Please cite this article in press as: Karakostas, A., Zizzo, D.J., Compliance and the power of authority. J. Econ. Behav. Organ. (2015), http://dx.doi.org/10.1016/j.jebo.2015.09.016